

Fig.1

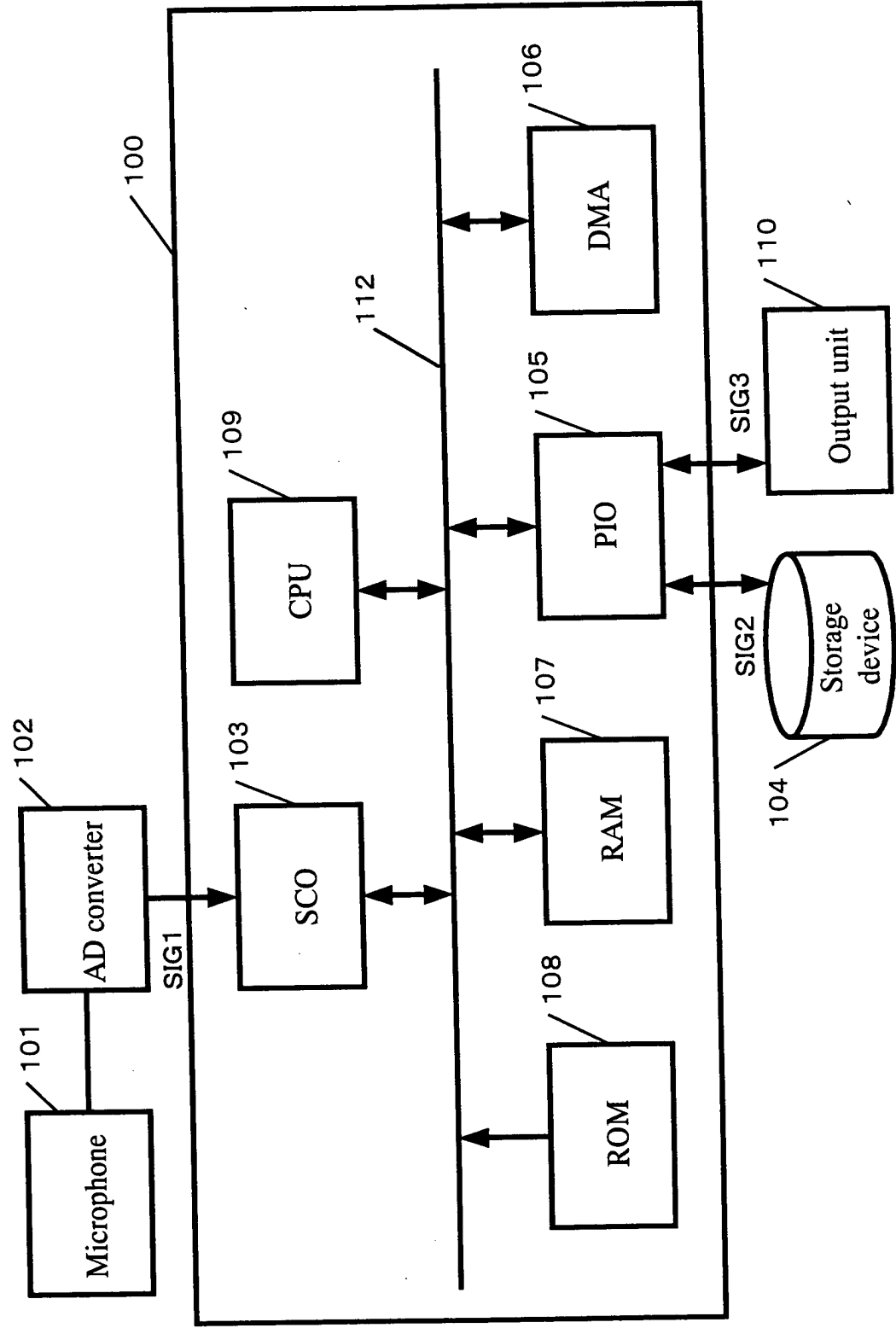


Fig.2

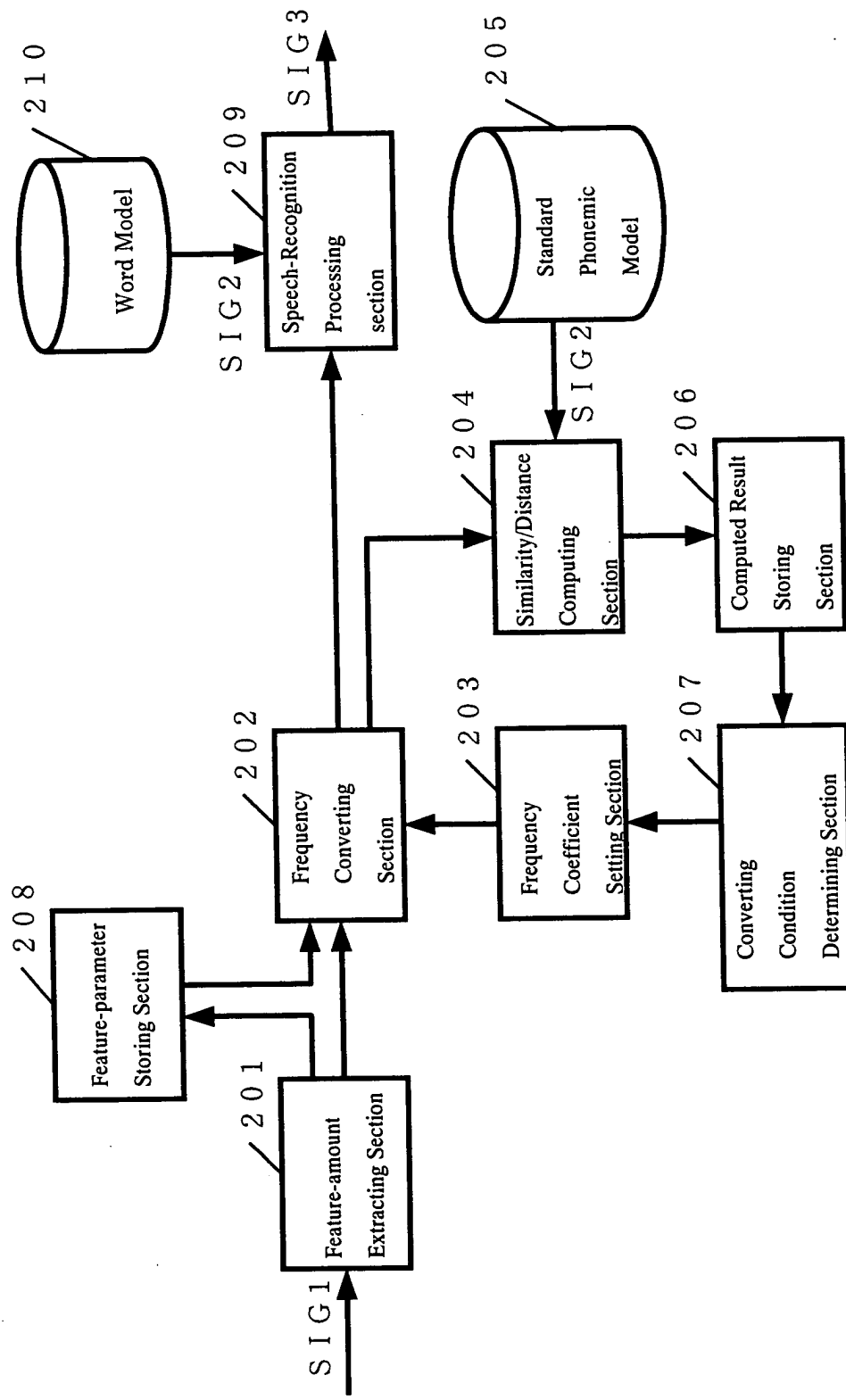


Fig.3

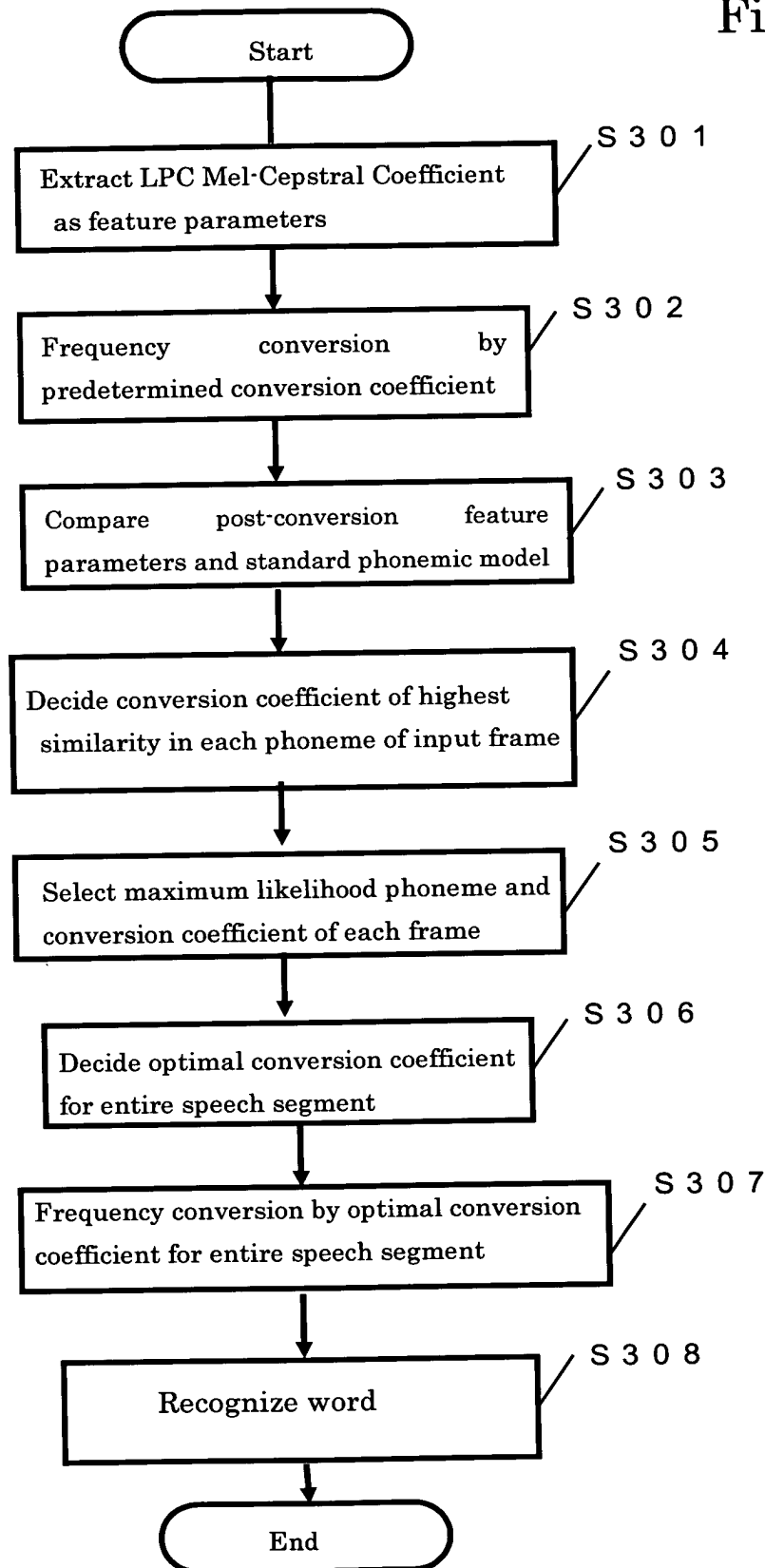


Fig.4

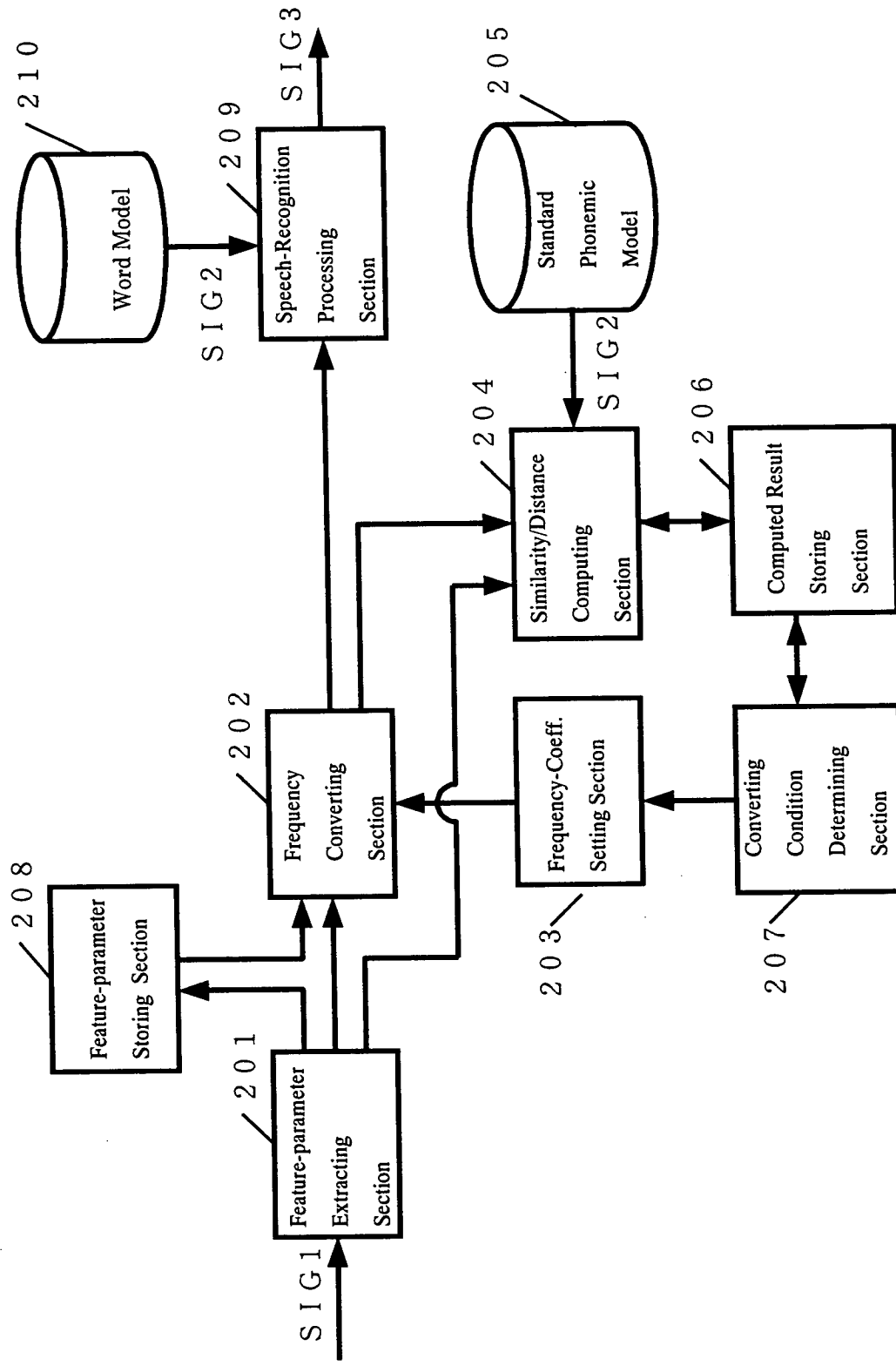


Fig.5

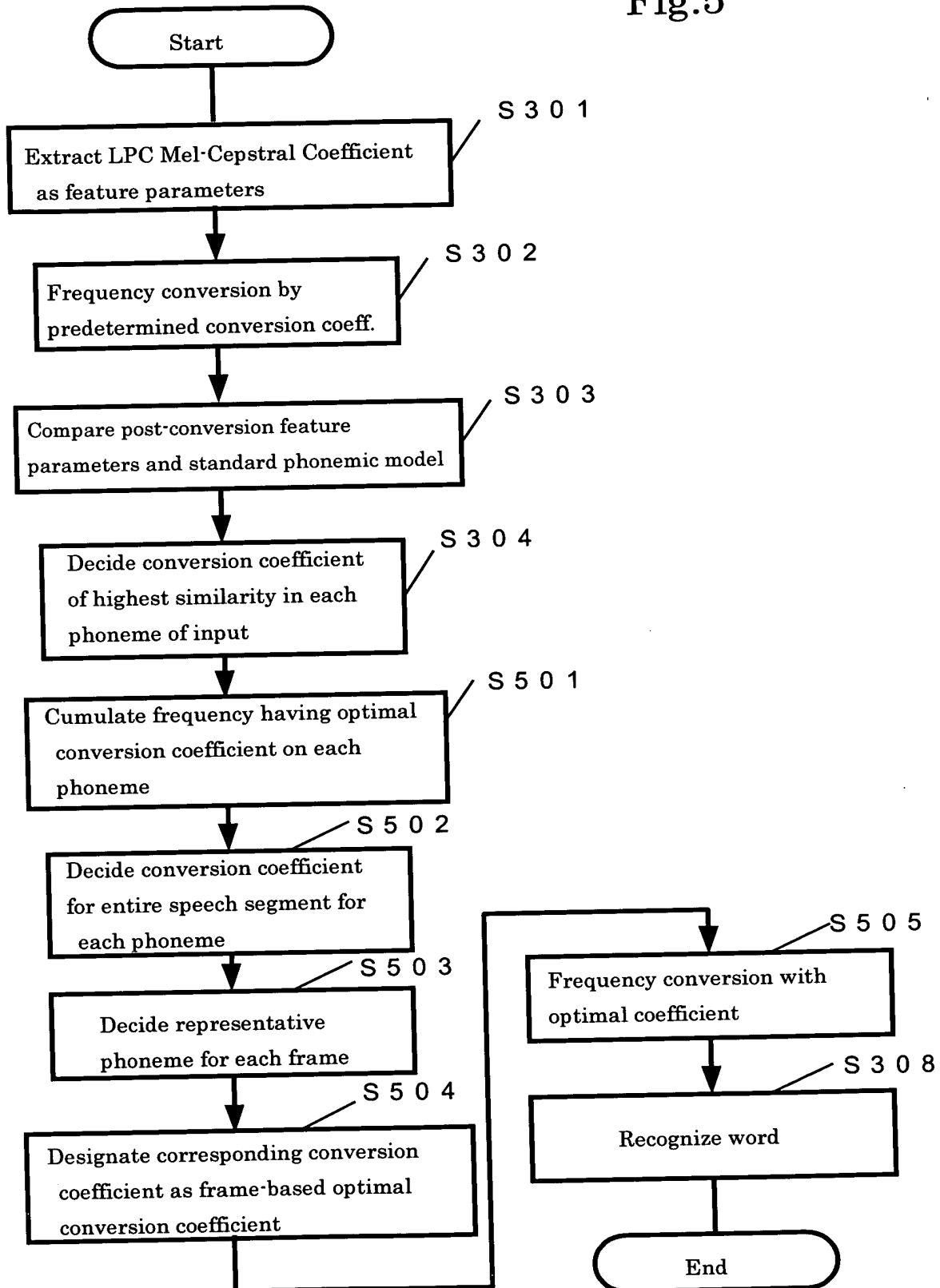


Fig.6

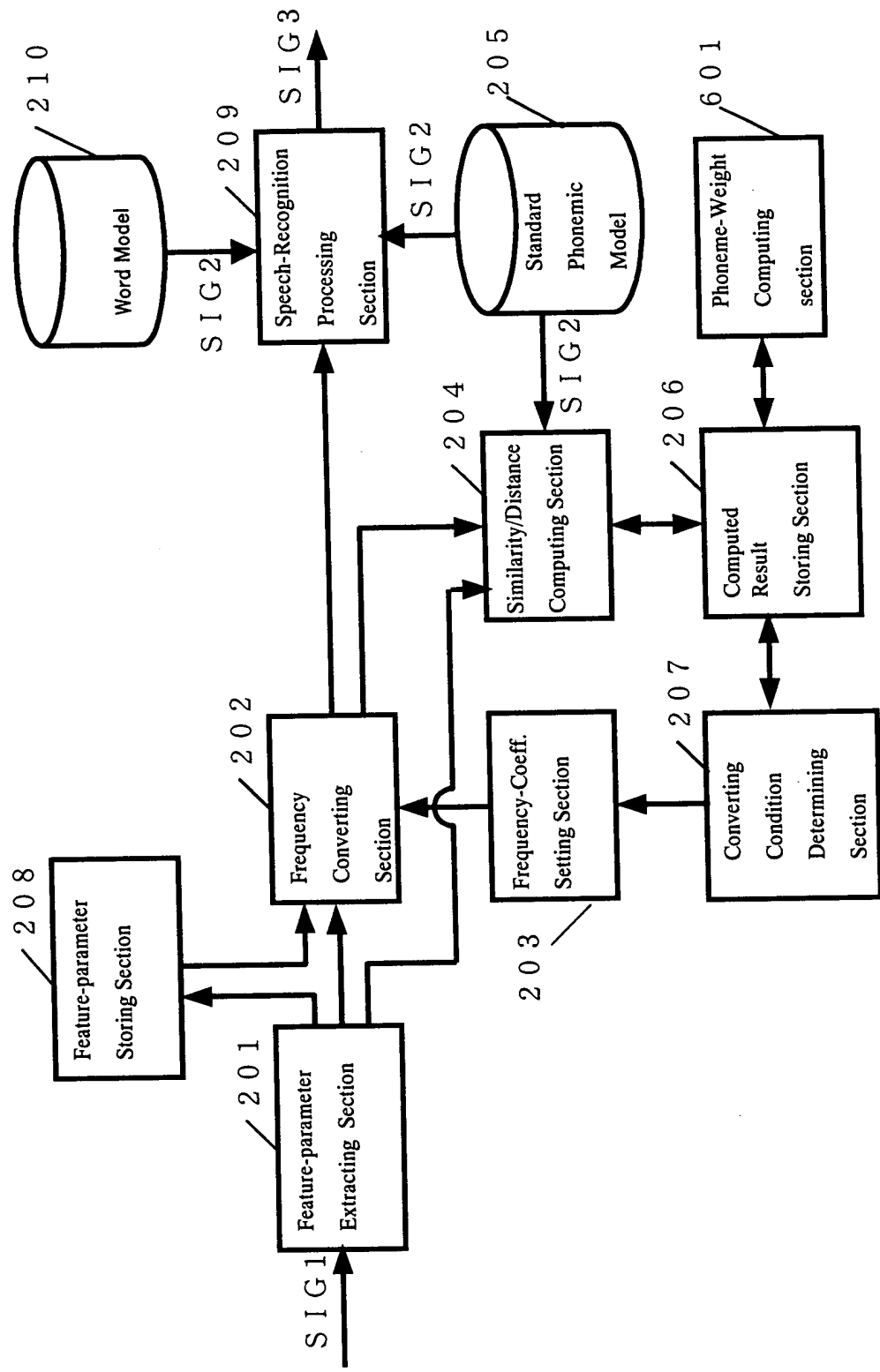


Fig.7

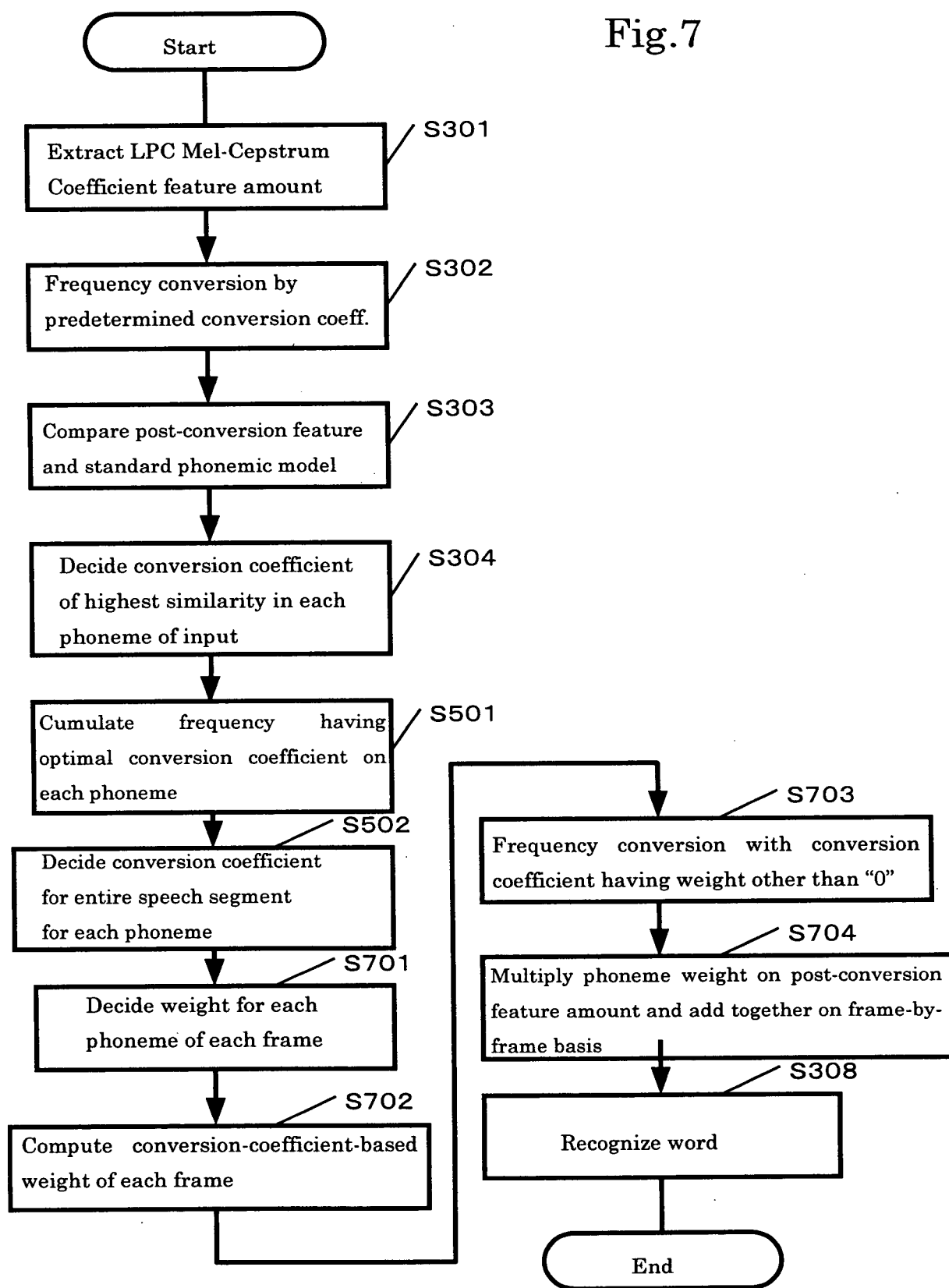


Fig.8A

801					802	803
Frame No.	/a/	/e/	/i/ /n/	Maximum likelihood of phoneme	Corresponding conversion coefficient
	α_4	α_3	α_5		/a/	α_4
First frame					/i/	α_3
Second frame	α_2	α_2	α_3			
.				.	.	.
.				.	.	.
.				.	.	.

Fig.8B

Conversion coefficient	Number of times
α_1	0
α_2	5
α_3	48
α_4	112
α_5	21
α_6	3
α_7	0

Fig.9A

Phoneme	α_1	α_2	α_3	α_4	α_5	α_6	α_7	Conversion coefficient where cumulative count is maximum
/a/	0	0	8	26	12	0	0	α_4
/e/	0	3	39	4	0	0	0	α_3
.				.				.
.				.				.
.				.				.

Fig.9B

Frame No.	Selected Phoneme	Conversion coefficient where cumulative count is maximum
First frame	/a/	α_4
Second frame	/k/	α_4
.	.	.
.	.	.
i-th frame	/n/	α_3
.	.	.
.	.	.

Fig.10A

Frame No.	/a/	/e/	/i/	/n/
First frame	w_{1a}	w_{1e}	w_{1i}	w_{1n}
Second frame	w_{2a}	w_{2e}	w_{2i}	w_{2n}
.				.		
.				.		
.				.		

Fig.10B

Frame No.	α_1	α_2	α_3
First frame	$w_{1a} + w_{1u}$	w_{1i}	$w_{1p} + w_{1e} +$
Second frame	$w_{2k} + w_{2u}$	w_{2a}	$w_{2e} + w_{2r} +$
.			.	
.			.	
.			.	

Fig.11A

Speaker	With speaker normalization (%)	Without speaker normalization (%)
Speaker A (boy)	8 9	8 2
Speaker B (boy)	8 3	7 5
Speaker C (girl)	9 6	7 5

Fig.11B

Speaker	With speaker normalization (%)	Without speaker normalization (%)
Adult men (two persons average)	9 9	9 9
Adult women (Three persons average)	9 9 . 3	9 8 . 7
Children (four persons average)	9 2	8 3 . 8

Fig.11C

Speaker	With speaker normalization (%)	Without speaker normalization (%)
Adult men (two persons average)	9 9	9 9
Adult women (Three persons average)	9 9 . 3	9 8 . 7
Children (four persons average)	9 3	8 3 . 8

Fig.12

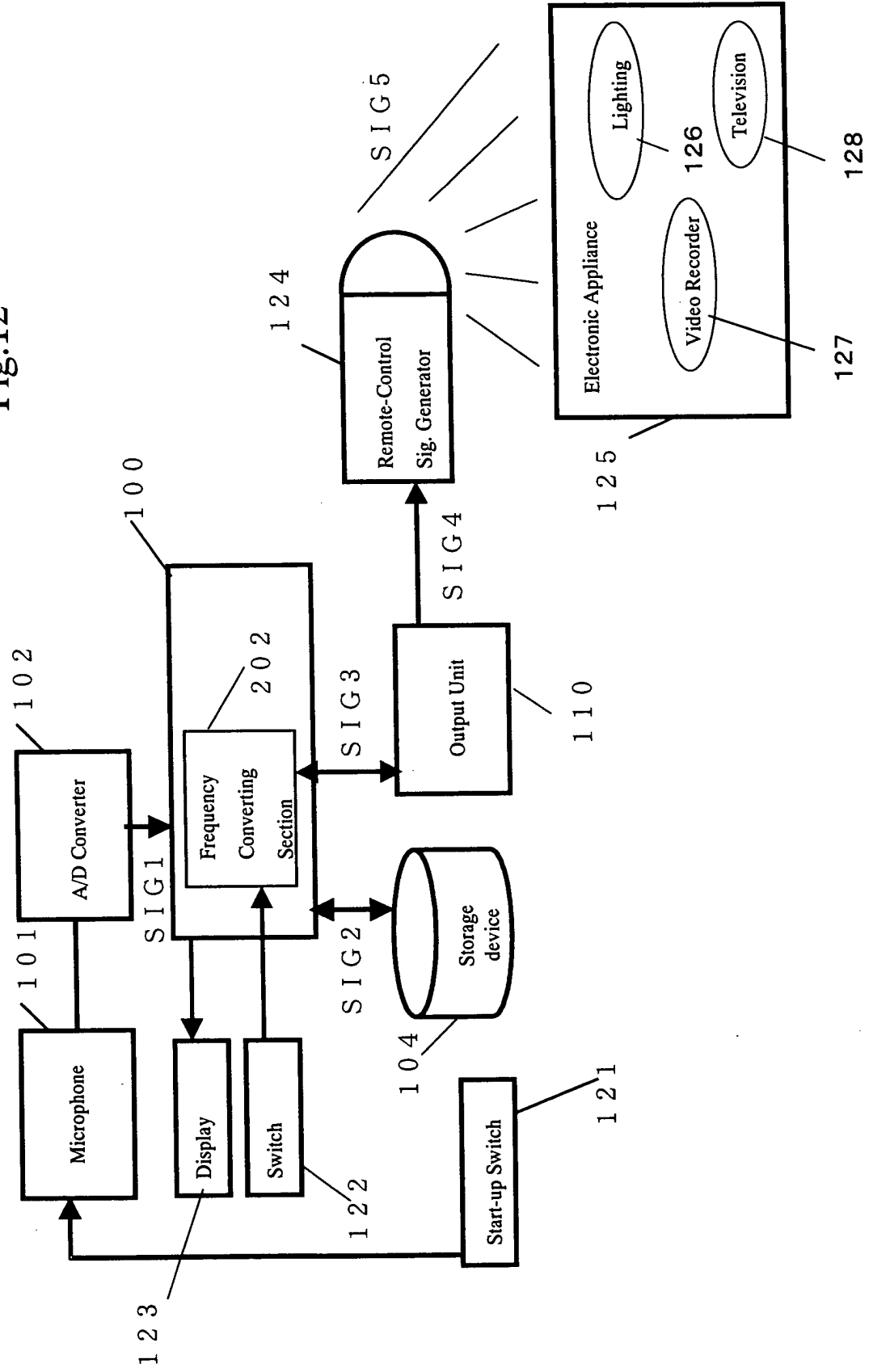


Fig.13

